JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **205,** 471–484 (1997)
ARTICLE NO. AY975213

Transformations of Lebesgue–Stieltjes Integrals*

Benjamin B. Winter†

Department of Mathematics and Statistics, University of Ottawa, P.O. Box 450, Station A, Ottawa, Ontario, K1N 6N5, Canada

Received May 2, 1995

INTRODUCTION

The Lebesgue–Stieltjes integral $\int f\,dG$ plays a central role in many branches of applied mathematics—for instance, in stochastic process [4, 6]. The use of such integrals is frequently facilitated by appropriate transformations. This note examines "change of variable" transformations, where an integral with integrand $f \circ h$ is transformed into one with integrand f, as well as the transformation of an integral with integrator $\varphi \circ G$ into an integral with respect to G.

In his discussion of the Stieltjes integral, Lebesgue [5] starts with the definition of $\int_a^b f(x)\,d\alpha(x)$ when f is continuous and α is of bounded variation and then, on p. 256, does a "change of variable" for the case where α is continuous and strictly increasing. As he puts it,

> the change of variable $\alpha = \alpha(x)$ transforms $f(x)$ into a function $g(\alpha)$ and transforms the definition of $\int f(x)\,d\alpha(x)$ into that of an ordinary integral of $g(\alpha)$, so that $\int_a^b f(x)\,d\alpha(x) = \int_{\alpha(a)}^{\alpha(b)} g(\alpha)\,d\alpha$ and one is brought back to ordinary integration of a continuous function.

What does the stated equality mean? (What is g? Does α have the same meaning throughout the stated equality?) Can it be proved rigorously? (Surely Lebesgue's minimalist argument does not constitute a proof.) Can it be generalized?

* The preparation of this article was supported in part by the Natural Sciences and Engineering Research Council Canada.

† E-mail address: bwinter@sciences.uottawa.ca.

0022-247X/97 $25.00

To see what Lebesgue had in mind, subject the initial integral to a purely formal manipulation (as in elementary calculus): "put $y = \alpha(x)$, so that $x = \alpha^{-1}(y)$, substitute for x, and adjust the limits of integration." This recipe transforms the expression $\int_a^b f(x)\, d\alpha(x)$ into $\int_{\alpha(a)}^{\alpha(b)} f(\alpha^{-1}(y))\, dy$. Comparing this with the equality stated by Lebesgue we see that, in that equality:

- g is the composition $f \circ \alpha^{-1}$;
- in "$g(\alpha)\, d\alpha$," α is a "variable of integration" and could be replaced by any other letter;
- elsewhere—in the limits of integration on the right side and in "$d\alpha$" on the left side—α is the integrator function.

Thus it appears that Lebesgue was asserting that $\int_a^b f(x)\, d\alpha(x) = \int_{\alpha(a)}^{\alpha(b)} f(\alpha^{-1}(y))\, dy$, when f is continuous and (to ensure the existence of the inverse) α is continuous and strictly increasing.

Lebesgue's result will be generalized in Theorem (2) to the case where f is measurable and α is a right-continuous function which is increasing but need not be strictly increasing (and the inverse α^{-1} is replaced by a suitable generalization). Corollary (3) will then show that the change-of-variable manipulations of elementary calculus are valid for the Lebesgue–Stieltjes integral with respect to a continuous integrator.

The results in Theorem (2) and Corollary (3) deal with change-of-variable results. Further results, in Theorems (5) and (6), deal with the transformation of an integral with integrator $\varphi \circ G$ into an integral with respect to G, when φ is increasing. In a suggestive shorthand, Theorem (5) shows that $d(\varphi \circ G) = \varphi'(G)\, dG$ when G is continuous and φ is everywhere differentiable. The results in Theorem (6) are for the case where G is a discrete distribution function.

Notation and Terminology

The σ-algebras of Borel sets of $\mathbb{R}$ and $\overline{\mathbb{R}}$ (the extended real line) are denoted, respectively, $\mathscr{B}$ and $\overline{\mathscr{B}}$. The notation $f\colon (\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ indicates that f is measurable function from $(\mathbb{R}, \mathscr{B})$ to $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$; i.e., f is a function from $\mathbb{R}$ to $\overline{\mathbb{R}}$, measurable with respect to the indicated σ-algebras. For a set A, $\mathbf{I}_A$ is the indicator (or characteristic) function of A. Set-theoretic difference is noted $A - B$.

If G is $\mathbb{R} \to \mathbb{R}$ then $G(-\infty), G(x^-), G(x^+), G(\infty)$ denote the obvious limits and, for $B \subset \mathbb{R}, G^{-1}(B) := \{x \in \mathbb{R} : G(x) \in B\}$. If $G\colon\ \mathbb{R} \to \mathbb{R}$ is increasing then $G_-(x) := G(x^-)$, $\mathscr{I}_G$ denotes the interval $(G(-\infty), G(\infty))$, and, for $y \in \mathbb{R}$, $G^\vee(y) := \inf\{x \in \mathbb{R}\colon G(x) \geq y\}$. This defines an increasing function on $\mathbb{R}$, with values in $\overline{\mathbb{R}}$, with $G^\vee(y) = -\infty$ if $y \leq G(-\infty)$ and $G^\vee(y) = \infty$ if $y > G(\infty)$.

A *distribution function* [d.f.] is a function from $\mathbb{R}$ to $\mathbb{R}$, increasing (i.e., nondecreasing) and right-continuous. If G is a d.f., μ_G denotes the *measure determined by* G, namely the unique measure on $\mathscr{B}$ which gives mass $G(b) - G(a)$ to an interval $(a, b]$, and the *Lebesgue–Stieltjes integral* $\int \cdot\, dG$ is the Lebesgue integral with respect to μ_G. Lebesgue measure (on $\mathscr{B}$) is denoted λ and $\int \cdot\, dx$ denotes integration with respect to λ. For equalities involving integrals, saying that *expression* 1 = *expression* 2 *in the usual sense* means that if either expression exists then both do and they are equal.

INVERSES OF INCREASING FUNCTIONS

In proofs and statements of results, this note involves inverses of distribution functions. If $G\colon \mathbb{R} \to \mathbb{R}$ is continuous and strictly increasing, it has an inverse. If the increasing function G has discontinuities or is not strictly increasing then, strictly speaking, it has no inverse function. Instead of a strict-sense inverse, one can rely on the mapping $G^{\vee}$ defined above, or the mapping $G^{\sim}\colon y \mapsto \inf\{x \in \mathbb{R}\colon G(x) > y\}$. By a common abuse of language, these are also called inverses (the left-continuous and right-continuous inverse, respectively). This view of the "inverse" is common in recent writings but some older treatises takes a different approach. For instance, Riesz and Sz.-Nagy [8] and Kamke [3] write $G^{-1}(y)$ for the set $\{x\colon G(x^-) \le y \le G(x^+)\}$ and, departing from the usual notation for functions, they use the symbol $G(x)$ to denote the set $\{y\colon G(x^-) \le y \le G(x^+)\}$.

Properties of $G^{\vee}$ needed in this note are set out in Proposition (1). They are similar to properties stated on pp. 191 and 200 of Billingsley [1], where G is a probability distribution function, and on p. 108 of Kopp [4], where it is assumed that G is right-continuous, with values in $[0, \infty)$ and $G(0^-) = 0$. Many useful properties of the left-continuous inverse of a probability distribution function are given in Parzen [7], a study in which that inverse plays a central role. A detailed study of left- and right-continuous inverses of increasing functions from $\mathbb{R}$ to $\overline{\mathbb{R}}$ can be found in Winter [12].

(1) PROPOSITION. *Let G be a distribution function.*

(a) $G^{\vee}(y)$ *is finite when* $y \in \mathscr{I}_G$.

(b) $G_-(G^{\vee}(y)) \le y \le G(G^{\vee}(y))$ *when* $G^{\vee}(y)$ *is finite*.

(c) $\{x \in \mathbb{R}\colon G(x) \ge y\} = [G^{\vee}(y), \infty)$ *when* $G^{\vee}(y)$ *is finite*.

(d) $a < G^{\vee}(y) \le b$ *iff* $G(a) < y \le G(b)$, *when* $G^{\vee}(y)$ *is finite and* $-\infty < a < b < \infty$.

(e) $G^{\vee}(c) \le x < G^{\vee}(d)$ *iff* $c \le G(x) < d$, *when* $G(-\infty) < c < d < G(\infty)$ *and* $x \in \mathbb{R}$.

Put $\Gamma_y := \{x \in \mathbb{R}\colon G(x) \geq y\}$, so that $G^{\vee}(y) = \inf \Gamma_y$. Now the above properties can be established by straightforward manipulation and careful examination of Γ_y.

CHANGE OF VARIABLE

Part (a) of Theorem (2) deals with $\int_{G(a)}^{G(b)} f(G^{\vee}(y))\,dy$, for $f\colon \mathbb{R} \to \overline{\mathbb{R}}$ and a d.f. G. Since $G^{\vee}(y)$ is infinite when $y \leq G(-\infty)$ or $y > G(\infty)$, the expression $f(G^{\vee}(y))$ is not necessarily defined for every $y \in \mathbb{R}$; however, by (1), it is defined when $y \in (G(a), G(b))$. Therefore the right side of (a) in Theorem (2) should be interpreted as $\int_{G(a)}^{G(b)} \varphi(y)\,dy$, where $\varphi(y)$ is $f(G^{\vee}(y))$ or 0, according to whether $G^{\vee}(y)$ is or is not finite. Analogous comments apply to part (b) of the theorem.

(2) THEOREM. *Consider a distribution function G, $f\colon (\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, and $-\infty < a < b < \infty$.*

(a) $\int_{(a,b]} f(x)\,dG(x) = \int_{G(a)}^{G(b)} f(G^{\vee}(y))\,dy$, *in the usual sense of such equalities.*

(b) *Statement (a) remains true if integration is over (a, ∞) on the left side and from $G(a)$ to $G(\infty)$ on the right, or over $(-\infty, b]$ on the left and from $G(-\infty)$ to $G(b)$ on the right, or over $(-\infty, \infty)$ on the left and from $G(-\infty)$ to $G(\infty)$ on the right.*

Proof. Consider the case where $f = \mathbf{I}_{(\alpha, \beta]}$. It is easily seen that if $(a, b] \cap (\alpha, \beta] = \varnothing$ then both sides of (a) equal 0. Suppose therefore that $(a, b] \cap (\alpha, \beta] \neq \varnothing$, so that $(a, b] \cap (\alpha, \beta] = (s, t]$ with $s = a \vee \alpha$ (i.e., $\max\{a, \alpha\}$) and $t = b \wedge \beta$ (i.e., $\min\{b, \beta\}$). The left side of (a) equals $G(t) - G(s)$. By Proposition (1), if $y \in (G(a), G(b))$ then $(\alpha < G^{\vee}(y) \leq \beta) \Leftrightarrow (G(\alpha) < y \leq G(\beta))$. Thus the right side of (a) equals

$$
\begin{aligned}
&\int_{\mathbb{R}} \mathbf{I}_{(G(a), G(b))}(y) \mathbf{I}_{(\alpha, \beta]}(G^{\vee}(y))\,dy \\
&\qquad = \int_{\mathbb{R}} \mathbf{I}_{(G(a), G(b)]}(y) \mathbf{I}_{(G(\alpha), G(\beta)]}(y)\,dy \\
&\qquad = G(b) \wedge G(\beta) - G(a) \vee G(\alpha) \\
&\qquad = G(b \wedge \beta) - G(a \vee \alpha) = G(t) - G(s).
\end{aligned}
$$

Thus (a) is true when f is the indicator of an interval $(\alpha, \beta]$.

For $B \in \mathscr{B}$, put

$$\mu_1(B) := \int_{(a,b]} \mathbf{I}_B(x)\, dG(x) \qquad \text{and}$$

$$\mu_2(B) := \int_{(G(a),G(b)]} \mathbf{I}_B\big(G^{\vee}(y)\big)\, dy$$

so that $\mu_1(B) = \mu_2(B)$ whenever B is an interval $(\alpha, \beta]$. Observe that if n is so large that $(a,b] \subset (-n,n]$ then $\mu_2(-n,n] = \mu_1(-n,n] = G(b) - G(a)$ is finite. Therefore (see, e.g., Theorem 10.3 in [1]) μ_1 and μ_2 agree on $\mathscr{B}$; i.e., (a) is true whenever f is the indicator of a Borel set. A standard argument then shows that (a) is true for any nonnegative f: $(\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$. When f is not necessarily nonnegative, apply what was just shown to f^+ and f^-.

For (b), use (a) and the monotone convergence theorem to obtain the desired conclusion for nonnegative f and then consider f^+ and f^-. ∎

The result in Theorem (2)(a), stated here for the case where G is an arbitrary d.f., is similar to a result stated on p. 13 of [6] where G is assumed to be a probability distribution function. The result in Theorem (2)(b) generalizes a result on p. 108 of [4], where the domain of G and f is $[0, \infty)$ rather than all of $\mathbb{R}$.

The results in Theorem (2) are obtained by elementary methods. A much more elaborate treatment yields somewhat more general results. For instance, an argument leading to a result similar to Theorem (2)(a) is outlined by Riesz and Sz.-Nagy [8, pp. 124–125]; they treat the inverse somewhat differently—see the comment just before Proposition (1)—and obtain a result where G is of bounded variation, without being required to be a distribution function. A similar generalization (which also allows for integration over Borel sets, and not just over intervals) is stated on p. 164 of [3].

(3) COROLLARY. *Consider a continuous distribution function G, f: $(\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, and $-\infty < a < b < \infty$.*

(a) $\int_{(a,b]} f(G(x))\, dG(x) = \int_{G(a)}^{G(b)} f(y)\, dy$, *in the usual sense of such equalities.*

(b) *Statement (a) remains true if integration is over (a, ∞) on the left side and from $G(a)$ to $G(\infty)$ on the right, or over $(-\infty, b]$ on the left and from $G(-\infty)$ to $G(b)$ on the right, or over $(-\infty, \infty)$ on the left and from $G(-\infty)$ to $G(\infty)$ on the right.*

Proof. Suppose that G is continuous. Now if $G(a) < y < G(b)$ then, by Proposition (1), $G(G^{\vee}(y)) = y$. Therefore, applying Theorem (2)(a) with f replaced by $f \circ G$,

$$\int_{(a,b]} f(G(x))\, dG(x) = \int_{G(a)}^{G(b)} f\big(G(G^{\vee}(y))\big)\, dy = \int_{G(a)}^{G(b)} f(y)\, dy.$$

Part (b) follows similarly from Theorem (2)(b). ∎

Corollary (3) shows that the "change of variable" manipulation of elementary calculus ("put $y = G(x)$ and $dy = dG(x)$, and adjust the limits of integration") gives a valid result when G is a continuous d.f. The result can break down when G is not continuous; for instance, if $f = G = \mathbf{I}_{[1,\infty)}$ then $f \circ G = f$ and $\int_{(0,2]} f(G(x))\, dG(x) = 1$ whereas $\int_{G(0)}^{G(2)} f(y)\, dy = 0$.

While the result in Corollary (3)(a) is essentially the same as the identity in problem 13 on p. 264 of Royden [9], the one-line proof given here is different from the argument suggested there. A generalization of Corollary (3)(a) is stated on pp. 164–165 of [3], showing that, under appropriate conditions, $\int_{(a,b]} f(G(x))\, d\varphi(G(x)) = \int_{G(a)}^{G(b)} f(y)\, d\varphi(y)$.

Theorem (2) has an interesting consequence which will be used in the proof of Theorem (5). It concerns the induced measure $\mu_G G^{-1}$ which assigns the value $\mu_G(G^{-1}(B))$ to $B \in \mathscr{B}$. That result, stated in the next proposition, generalizes problem 12.a on p. 264 of Royden [9] (and the proof given here is different from the argument suggested there).

(4) Proposition. *Consider a continuous distribution function G with range $\mathscr{R} = \{G(x)\colon x \in \mathbb{R}\}$ and let $|_{\mathscr{R}}$ denote the restriction of a measure to $\mathscr{R}$.*

(a) *For every $B \in \mathscr{B}$, $\mu_G(G^{-1}(B)) = \lambda(B \cap \mathscr{R})$.*

(b) $(\mu_G G^{-1})|_{\mathscr{R}} = \lambda|_{\mathscr{R}}$.

(c) *If $f\colon (\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ and A is a Borel subset of $\mathscr{R}$ then $\int_A f d(\mu_G G^{-1}) = \int_A f\, d\lambda$, in the usual sense of such equalities.*

Proof. From Theorem (2) and Proposition (1),

$$\begin{aligned}
\mu_G\big(G^{-1}(B)\big) &= \int_{\mathbb{R}} \mathbf{I}_{G^{-1}(B)}(x)\, dG(x) = \int_{\mathscr{I}_G} \mathbf{I}_{G^{-1}(B)}\big(G^{\vee}(y)\big)\, dy \\
&= \int_{\mathscr{I}_G} \mathbf{I}_B\big(G(G^{\vee}(y))\big)\, dy = \int_{\mathscr{I}_G} \mathbf{I}_B(y)\, dy \\
&= \lambda(B \cap \mathscr{I}_G) = \lambda(B \cap \mathscr{R}).
\end{aligned}$$

The last step is valid because, as G is monotone and continuous, if $\mathscr{R}$ is not equal to $\mathscr{I}_G$, it differs from it only by the addition of one or both of the points $G(-\infty)$ and $G(\infty)$. Part (b) follows because $G(x)$ is always an element of $\mathscr{R}$, so that $G^{-1}(B \cap \mathscr{R}) = G^{-1}(B)$. And (c) is an immediate consequence of (b). ∎

TRANSFORMATIONS OF CONTINUOUS DISTRIBUTION FUNCTIONS

(5) THEOREM. *Consider a continuous distribution function G and let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be increasing and differentiable at every $x \in \mathbb{R}$.*

(a) *For every $B \in \mathscr{B}$, $\mu_{\varphi \circ G}(B) = \int_B (\varphi' \circ G)\, dG$.*

(b) *If $f\colon (\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ and $A \in \mathscr{B}$ then $\int_A f\, d(\varphi \circ G) = \int_A f \cdot (\varphi' \circ G)\, dG$, in the usual sense of such equalities.*

Proof. Note that, as φ is increasing and differentiable at every $x \in \mathbb{R}$, φ' is a measurable function from $(\mathbb{R}, \mathscr{B})$ to $(\mathbb{R}, \mathscr{B})$ and $\int_a^b \varphi'(x)\, dx = \varphi(b) - \varphi(a)$; see, e.g., (18.14) in [2] and Theorem 7.21 in [10].

Write η for $\mu_{\varphi \circ G}$, the measure determined by $\varphi \circ G$. Note that $\varphi \circ G$ is a continuous d.f. and $\varphi' \circ G$ is a measurable function from $(\mathbb{R}, \mathscr{B})$ to $(\mathbb{R}, \mathscr{B})$.

Consider the nonempty bounded interval (c, d). If $G(c) = G(d)$ then

$$\int_{(c,d)} (\varphi' \circ G)\, dG = 0 \qquad \text{and} \qquad \eta(c,d) = \varphi(G(d)) - \varphi(G(c)) = 0.$$

It remains to consider the case $G(c) < G(d)$.

To begin with, suppose that $G(-\infty) < G(c) < G(d) < G(\infty)$. As $[G(c), G(d))$ is contained in the range of G, it follows from Proposition (4) that

$$\begin{aligned} \eta(c,d) = \varphi(G(d)) - \varphi(G(c)) &= \int_{[G(c), G(d))} \varphi'\, d\lambda \\ &= \int_{[G(c), G(d))} \varphi' d(\mu_G G^{-1}), \end{aligned} \tag{i}$$

where, as before, $\mu_G G^{-1}$ is the induced measure defined by $(\mu_G G^{-1})(B) := \mu_G(G^{-1}(B))$. Furthermore,

$$\int_{[G(c), G(d))} \varphi' d(\mu_G G^{-1}) = \int_{G^{-1}[G(c), G(d))} (\varphi' \circ G)\, d\mu_G; \tag{ii}$$

see, e.g., Theorem 16.12 in [1]. By Proposition (1),

$$\begin{aligned} G^{-1}[G(c), G(d)) &= \{x \in \mathbb{R}: G(c) \le G(x) < G(d)\} \\ &= \{x \in \mathbb{R}: G^{\vee}(G(c)) \le x < G^{\vee}(G(d))\}. \end{aligned}$$

Thus, from (i) and (ii),

$$\eta(c, d) = \int_{[\tilde{c}, \tilde{d})} (\varphi' \circ G)\, dG,$$

$$\text{where } \tilde{c} := G^{\vee}(G(c)) \text{ and } \tilde{d} := G^{\vee}(G(d)). \quad \text{(iii)}$$

Note that $G^{\vee}(G(c)) = \inf\{x \in \mathbb{R}: G(x) \ge G(c)\} \le c$; likewise, $G^{\vee}(G(d)) \le d$. Furthermore, $G^{\vee}(G(d)) = \inf\{x \in \mathbb{R}: G(x) \ge G(d)\} > c$, due to $G(c) < G(d)$ and right-continuity of G. Thus $\tilde{c} \le c < \tilde{d} \le d$. By Proposition (1) and continuity of G, $G(\tilde{c}) = G(G^{\vee}(G(c))) = G(c)$; likewise, $G(\tilde{d}) = G(d)$. Therefore, $\int_{[\tilde{c}, c)} (\varphi' \circ G)\, dG = \int_{[\tilde{d}, d)} (\varphi' \circ G)\, dG = 0$ and it follows from (iii) that

$$\eta(c, d) = \int_{(c, d)} (\varphi' \circ G)\, dG$$

$$\text{when } G(-\infty) < G(c) < G(d) < G(\infty). \quad \text{(iv)}$$

Now consider the general case, where $G(-\infty) \le G(c) < G(d) \le G(\infty)$. Put

$$\bar{c} := \begin{cases} c & \text{if } G(-\infty) < G(c) \\ \sup\{x \in \mathbb{R}: G(x) = G(-\infty)\} & \text{if } G(-\infty) = G(c) \end{cases}$$

and

$$\bar{d} := \begin{cases} d & \text{if } G(d) < G(\infty) \\ \inf\{x \in \mathbb{R}: G(x) = G(\infty)\} & \text{if } G(d) = G(\infty). \end{cases}$$

Examining the four possible combinations, one sees that $c \le \bar{c} < \bar{d} \le d$. Therefore there exist real sequences $(c_n)_1^{\infty}$ and $(d_n)_1^{\infty}$ such that $c_n \downarrow \bar{c}$, $d_n \uparrow \bar{d}$, and $G(-\infty) < G(c_n) < G(d_n) < G(\infty)$. By the monotone convergence theorem and continuity of G, it follows from (iv) that $\int_{(\bar{c}, \bar{d})} (\varphi' \circ G)\, dG = \eta(\bar{c}, \bar{d})$. Furthermore, since $\eta(c, \bar{c}] = 0 = \eta[\bar{d}, d)$ and $\mu_G(c, \bar{c}] = 0 = \mu_G[\bar{d}, d)$, it follows that $\eta(c, d) = \int_{(c, d)} (\varphi' \circ G)\, dG$.

If θ is the measure on $\mathscr{B}$ defined by $\theta(B) := \int_B (\varphi' \circ G)\, dG$, the above shows that η and θ agree on bounded open intervals. As these measures are σ-finite, it follows that they agree on $\mathscr{B}$; i.e., $\mu_{\varphi \circ G}(B) = \int_B (\varphi' \circ G)\, dG$ for every $B \in \mathscr{B}$. This completes the proof of (a), and (b) follows by a standard argument. ∎

In a suggestive shorthand, Theorem (5) states that $d(\varphi \circ G) = \varphi'(G)\,dG$ when G is a continuous d.f. and the increasing function φ is everywhere differentiable. For example, if φ is the exponential function and the d.f. G is continuous then, as one would expect, $\int_A f\,de^G = \int_A f \cdot e^G\,dG$ (in the usual sense)—but that is not necessarily true if G is discontinuous.

Both conclusions in Theorem (5) can fail to be true if G has discontinuities. For instance, if φ is the exponential function, $G = \mathbf{I}_{[0,\infty)}$, and $f \equiv 1$, then $\int_{\{0\}} f\,d(\varphi \circ G) = \int_{\{0\}} f\,de^G = e - 1$ while $\int_{\{0\}} f \cdot (\varphi' \circ G)\,dG = \int_{\{0\}} fe^G\,dG = e$, and $\mu_{\varphi \circ G}\{0\} = e - 1$ while $\int_{\{0\}}(\varphi' \circ G)\,dG = e$.

The conclusion in Theorem (5) can fail to be true if φ is not everywhere differentiable, even if φ is differentiable λ-a.e. or μ_G-a.e. For example, suppose that $G(x) = x$ and φ is the Lebesgue singular function on $[0,1]$ (see, e.g., (8.28) in [2] or 31.2 in [1]), also known as the Cantor function, extended so that $\varphi(x) = 0$ for $x < 0$ and $\varphi(x) = 1$ for $x > 1$. This φ is increasing and continuous, with $\varphi'(x) = 0$ λ-a.e., hence also μ_G-a.e., and with $\varphi(0) = 0$, $\varphi(1) = 1$. Now $\mu_{\varphi \circ G}[0,1] = 1$ while $\int_B(\varphi' \circ G)\,dG = 0$ for every $B \in \mathscr{B}$, and (b) fails with $f \equiv 1$ and $A = [0,1]$. [This example was suggested by C. M. Deo.]

Theorem (5) was obtained by elementary methods. It can also be obtained as an application of the following classical result on the differentiation of measures.

THEOREM. *Let σ and θ be finite measures on $\mathscr{B}_{\mathbb{X}}$, the σ-algebra of Borel sets of a separable metric space $\mathbb{X}$. For each $n \in \mathbb{N}^+$, let Π_n be a $\mathscr{B}_{\mathbb{X}}$-measurable countable partition of $\mathbb{X}$. Suppose that, for every $n \in \mathbb{N}^+$, every set in Π_{n+1} is contained in a set in Π_n, and that (with d denoting the diameter of a set) $\lim_{n\to\infty}\sup\{d(A)\colon A \in \Pi_n\} = 0$. For $x \in \mathbb{X}$, let M_n^x be the unique set in Π_n containing x and, with $0/0$ interpreted as $+\infty$, put*

$$\overline{D}(\sigma, \theta; x) := \limsup_{n\to\infty} \sigma(M_n^x)/\theta(M_n^x)$$

$$\underline{D}(\sigma, \theta; x) := \liminf_{n\to\infty} \sigma(M_n^x)/\theta(M_n^x)$$

$$E_\infty := \left\{x \in X\colon \underline{D}(\sigma, \theta; x) = \infty = \overline{D}(\sigma, \theta; x)\right\}.$$

Then

$$\underline{D}(\sigma, \theta; x) = \overline{D}(\sigma, \theta; x) \qquad \text{for } \theta\text{-almost all } x$$

and

$$\sigma(B) = \sigma(B \cap E_\infty) + \int_B \overline{D}(\sigma, \theta; x)\,d\theta(x) \qquad \text{for all } B \in \mathscr{B}_{\mathbb{X}}.$$

This is a slightly simplified statement of Theorem (15.7) in Chapter IV of [11]. With $-\infty < c < d < \infty$ and $\mathbb{X} = (c, d]$ with the usual metric, $\sigma(B) := \int_B \mathbf{I}_{(c,d]}\, d(\varphi \circ G)$, and $\theta(B) := \int_B \mathbf{I}_{(c,d]}\, dG$, one can use the above theorem to show that $\int_{\mathbb{R}} f\, d(\varphi \circ G) = \int_{\mathbb{R}} f \cdot (\varphi' \circ G)\, dG$ when $f = \mathbf{I}_{(c,d]}$; the general result then follows by a standard argument.

TRANSFORMATIONS OF DISCRETE DISTRIBUTION FUNCTIONS

Terminology. If μ is a measure on $\mathscr{B}$ which assigns finite mass to any bounded interval, a *distribution function of* μ if a function $G\colon \mathbb{R} \to \mathbb{R}$ such that $\mu(a, b] = G(b) - G(a)$ whenever $-\infty < a < b < \infty$. (Note that this defines *a*, and not *the*, distribution function of μ.) Every such function is indeed a distribution function—i.e., increasing and right-continuous. If θ is a measure on a σ-algebra $\mathscr{A}$ then saying that θ is *concentrated on* a set $C \in \mathscr{A}$ means that $\theta(B) = \theta(B \cap C)$ for every $B \in \mathscr{A}$ (equivalently, $\theta(B) = 0$ when $B \cap C = \varnothing$), and saying that θ is a *discrete measure* means that it is concentrated on some countable set. Saying that G is a *discrete distribution function* means that it is a distribution function of a discrete measure on $\mathscr{B}$. For a distribution function G, put $\Delta_G(x) := G(x) - G(x^-)$.

Generalized Sums. We will encounter expressions of the form $\sum_{x \in A} g(x)$ where g is a function from some domain D to $\overline{\mathbb{R}}$, nonnegative on $A \subset D$, and A may be an uncountable set. This can be defined, for instance, by putting $\sum_{x \in A} g(x) := \int_A g\, d\nu$ where ν is the counting measure on D. If the set $\{x \in A\colon g(x) > 0\}$ is countable and $(\xi_1, \xi_2, \ldots)$ is any enumeration of that set then $\sum_{x \in A} g(x) = \sum_{i \ge 1} g(\xi_i)$, this being a sum or a series, according to whether the set A is finite or infinite.

Preliminary Remarks. Consider distribution functions G and φ. Put $H := \varphi \circ G$ and $\theta(B) := \sum_{x \in B} \Delta_H(x)$, $B \in \mathscr{B}$.

- Clearly, H is also a d.f. It is easily seen that, if σ is any measure on $\mathscr{B}$, H is a d.f. of σ iff $\sigma = \mu_H$. Since Δ_H is a nonnegative function on $\mathbb{R}$ and generalized summation is just integration with respect to counting measure, the mapping $\theta\colon B \mapsto \sum_{x \in B} \Delta_H(x)$ is a measure on $\mathscr{B}$.

- Let G be a d.f. of a discrete measure which is concentrated on a finite set $\mathscr{M} \subset \mathbb{R}$, with elements $\xi_1 < \cdots < \xi_n$. In this case, $H(x^-) = \varphi(G(x^-))$ since G is constant on $(x - \delta, x)$ for sufficiently small $\delta > 0$. It is then easily seen that θ is concentrated on $\mathscr{M}$ and that H is a d.f. of θ. Then the measure determined by H is the measure θ; i.e.,

$$\mu_{\varphi \circ G}(B) = \textstyle\sum_{x \in B} \Delta_{\varphi \circ G}(x) \qquad \text{for every } B \in \mathscr{B}. \tag{i}$$

Furthermore, with $\nu =$ counting measure on $\mathbb{R}$, $\sum_{x \in B} \Delta_{\varphi \circ G}(x) = \int_B \Delta_{\varphi \circ G}\, d\nu$, so that $\mu_{\varphi \circ G}(B) = \int_B \Delta_{\varphi \circ G}\, d\nu$. Consequently, if $f\colon (\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ then, in the usual sense,

$$\int_A f d(\varphi \circ G) = \int_A f \cdot \Delta_{\varphi \circ G}\, d\nu = \sum_{x \in A} f(x)\{\varphi(G(x)) - \varphi(G(x^-))\}.$$

Thus (in the usual sense)

$$\int_A f d(\varphi \circ G) = \sum_{\xi_i \in A} f(\xi_i)\{\varphi(G(\xi_i)) - \varphi(G(\xi_i^-))\}. \tag{ii}$$

One might be inclined to believe that (i) and (ii) are true when G is any discrete d.f. It will be shown that these identities are indeed true when G is discrete and φ is absolutely continuous but that they may fail to be true when φ is merely continuous.

• More generally, let G be a d.f. of a discrete measure concentrated on a countable set $\mathscr{M}$; that set may be infinite and may have points of accumulation. Suppose that φ is continuous. Now $H(x^-) = \varphi(G(x^-))$. [Unlike the case where $\mathscr{M}$ is finite, this equality may fail to be true—even though G is discrete—when φ is not continuous.] Therefore $\Delta_H(x) = 0$ if $x \notin \mathscr{M}$, hence $(B \cap \mathscr{M} = \varnothing) \Rightarrow (\theta(B) = \sum_{x \in B} \Delta_H(x) = 0)$. As $\mathscr{M}$ is countable, we see that θ is a discrete measure, concentrated on $\mathscr{M}$. Now, since $\theta(c, d] = \sum_{x \in (c,d]} \Delta_H(x)$, it would appear plausible that, as in the case where $\mathscr{M}$ is finite, H is a d.f. of θ, i.e., that $\theta(c, d]$ in fact equals $H(d) - H(c)$. It will be seen that this is indeed the case when φ is absolutely continuous, but may fail to be true if φ is merely continuous.

(6) THEOREM. *Consider G, a distribution function of a discrete measure that is defined on $\mathscr{B}$ and concentrated on a countable set $\mathscr{M}$, and let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be increasing and absolutely continuous on every compact interval.*

(a) *Put $\theta(B) := \sum_{x \in B} \Delta_{\varphi \circ G}(x)$, $B \in \mathscr{B}$. Then θ is a discrete measure, concentrated on $\mathscr{M}$, $\varphi \circ G$ is a distribution function of θ, and $\theta = \mu_{\varphi \circ G}$.*

(b) *If $f\colon (\mathbb{R}, \mathscr{B}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ and $A \in \mathscr{B}$ then, in the usual sense of such equalities,*

$$\begin{aligned}\int_A f d(\varphi \circ G) &= \sum_{x \in A} f(x) \Delta_{\varphi \circ G}(x) \\ &= \sum_{x \in A \cap \mathscr{M}} f(x)\{\varphi(G(x)) - \varphi(G(x^-))\}.\end{aligned}$$

Proof. The fact that θ is discrete and concentrated on $\mathscr{M}$ was already noted in the preliminary remarks. To show that $H \coloneqq \varphi \circ G$ is a d.f. of θ, let γ be a measure with d.f. G and consider arbitrary but fixed $-\infty < c < d < \infty$.

If $(c, d] \cap \mathscr{M} = \varnothing$ then $\theta(c, d] = 0$ and $G(d) - G(c) = \gamma(c, d] = 0$, hence $H(d) = \varphi(G(d)) = \varphi(G(c)) = H(c)$, so that $H(d) - H(c) = 0 = \theta(c, d]$ is this case.

If $(c, d]$ contains n points of $\mathscr{M}$, say $\xi_1 < \cdots < \xi_n$, then these are the only points x in $(c, d]$ where $\Delta_H(x) > 0$ is possible. Therefore

$$\theta(c, d] = \Delta_H(\xi_1) + \cdots + \Delta_H(\xi_n) = H(\xi_n^+) - H(\xi_1^-) = H(d) - H(c).$$

Now consider the remaining case, where $(c, d] \cap \mathscr{M}$ is countably infinite. Let $(\xi_1, \xi_2, \dots)$ be an enumeration of $(c, d] \cap \mathscr{M}$ and put

$$\Gamma_i \coloneqq \big(G(\xi_i^-), G(\xi_i)\big] \qquad \text{and} \qquad \Gamma \coloneqq \bigcup_{i=1}^{\infty} \Gamma_i.$$

Since $c < \xi_i \leq d$ implies $G(c) \leq G(\xi_i^-) \leq G(\xi_i) \leq G(d)$, it follows that $\Gamma \subset (G(c), G(d)]$. Since $\xi_i < \xi_j$ implies $G(\xi_i) \leq G(\xi_j^-)$, it follows that Γ_i and Γ_j are disjoint when $i \neq j$. Therefore

$$\lambda(\Gamma) = \sum_{i=1}^{\infty} \lambda(\Gamma_i) = \sum_{i=1}^{\infty} \{G(\xi_i) - G(\xi_i^-)\} = \gamma(c, d] = G(d) - G(c),$$

so that $\lambda((G(c), G(d)] - \Gamma) = \lambda(G(c), G(d)] - \lambda(\Gamma) = 0$. Since φ is monotone, φ' exists λ-a.e.; since φ is absolutely continuous on $[G(c), G(d)]$,

$$\begin{aligned} H(d) - H(c) &= \varphi(G(d)) - \varphi(G(c)) = \int_{(G(c), G(d)]} \varphi'(x)\, dx \\ &= \int_{(G(c), G(d)] \cap \Gamma} \varphi'(x)\, dx + \int_{(G(c), G(d)] - \Gamma} \varphi'(x)\, dx. \end{aligned}$$

Since $\lambda((G(c), G(d)] - \Gamma) = 0$, the last integral equals zero. Furthermore, since $H(x^-) = \varphi(G(x^-))$, the next-to-last integral equals

$$\begin{aligned} \int_\Gamma \varphi'\, d\lambda &= \sum_{i=1}^{\infty} \int_{\Gamma_i} \varphi'\, d\lambda = \sum_{i=1}^{\infty} \{\varphi(G(\xi_i)) - \varphi(G(\xi_i^-))\} \\ &= \sum_{i=1}^{\infty} \{H(\xi_i) - H(\xi_i^-)\} = \sum_{x \in (c, d]} \Delta_H(x) = \theta(c, d]. \end{aligned}$$

Therefore, as in the first two cases, $H(d) - H(c) = \theta(c, d]$.

Thus $H = \varphi \circ G$ is a d.f. of θ and (see the preliminary remarks) it follows that $\theta = \mu_H$; i.e., for $B \in \mathscr{B}$,

$$\mu_{\varphi \circ G}(B) = \sum_{x \in B} \Delta_{\varphi \circ G}(x) = \int_B \Delta_{\varphi \circ G}\, d\nu,$$

where $\nu =$ counting measure on $\mathbb{R}$. It follows that (in the usual sense)

$$\int_A f\,d(\varphi \circ G) = \int_A f \Delta_{\varphi \circ G}\,d\nu$$
$$= \sum_{x \in A} f(x) \Delta_{\varphi \circ G}(x) = \sum_{x \in A} f(x)\{\varphi(G(x)) - \varphi(G(x^-))\}$$

and the last general summation can in fact be taken over $A \cap \mathscr{M}$. ∎

DISCUSSION

If G is a d.f. of a discrete measure and $\varphi: \mathbb{R} \to \mathbb{R}$ is continuous, but not absolutely continuous, then $H := \varphi \circ G$ may fail to be a d.f. of the measure θ defined by $\theta(B) := \sum_{x \in B} \Delta_H(x)$. For example, as in the discussion following Theorem (5), let φ be the Lebesgue singular function, extended so that $\varphi(x) = 0$ for $x < 0$ and $\varphi(x) = 1$ for $x > 1$. This φ is increasing and continuous but not absolutely continuous. Put $G(y) := \varphi^{\vee}(y)$, $0 < y < 1$, and $G(y) := 0$ for $y \leq 0$, $G(y) := 1$ for $y \geq 1$. Then G is a d.f. of the discrete probability measure γ which, for every $n \in \mathbb{N}^+$ and $1 \leq k \leq 2^{n-1}$, puts mass $1/3^n$ at the point $\xi_{nk} = (2k-1)/2^n$. At stage n of the aglorithm for constructing the Cantor ternary set, one removes 2^{n-1} intervals of length 3^{-n} from the set remaining at stage $n-1$; let these intervals be labeled $I_{n,k}$ ($k = 1, \dots, 2^{n-1}$), the numbering being such that $I_{n,1}$ is the "leftmost" interval and, going from left to right, $I_{n,k}$ is followed by $I_{n,k+1}$. Then $G(\xi_{nk})$ is the right end of $I_{n,k}$ and $G(\xi_{nk}^-)$ is the left end of that interval; as φ is constant on every one of the "middle third" intervals, it follows that $\varphi(G(\xi_{nk})) = \varphi(G(\xi_{nk}^-))$, hence $\Delta H(\xi_{nk}) = 0$. Thus θ is the null measure and, since $\varphi(0) = 0 = G(0)$ and $\varphi(1) = 1 = G(1)$, so that $H(0) = 0$ and $H(1) = 1$, it is clear that H is not a d.f. of θ. [This example was suggested by C. M. Deo.]

With G, φ, and H as above, it is not true that $\mu_{\varphi \circ G}(B) = \sum_{x \in B} \Delta_{\varphi \circ G}(x)$ for all $B \in \mathscr{B}$. As noted above, the right side is identically 0; on the other hand, the left side equals 1 when $B = [0,1]$. Also, it is not true that $\int_A f\,d(\varphi \circ G) = \sum_{x \in A \cap \mathscr{M}} f(x)\{\varphi(G(x)) - \varphi(G(x^-))\}$; the right side is always 0, whereas the left side equals 1 when $A = [0,1]$ and $f \equiv 1$.

In Theorem (6), it is assumed that φ is absolutely continuous on every compact interval, and not that φ is absolutely continuous (on $\mathbb{R}$). The former condition is weaker than the latter. For example, let φ be a function which vanishes on $(-\infty, 0]$ and whose graph (to the right of the origin) is obtained by joining by straight-line segments the points $p_0, p_1, \dots$ where $p_0 = (0,0)$ and, for $k \in \mathbb{N}^+$, $p_k = (k, 1 + 2 + \cdots + k)$. Now φ is absolutely continuous on any compact interval $[a,b]$ because, for $x \in$

$[a, b]$, $\varphi(x) = \int_a^x \varphi'(s)\,ds$. But φ is not absolutely continuous on $\mathbb{R}$ because, given $\epsilon > 0$, one can, for any $\delta > 0$, find x so large that $\varphi(x + \delta/2) - \varphi(x) > \epsilon$.

Theorem (6) asserts a relationship between distribution functions, and *not* between measures. More explicitly, let G be a d.f. of a discrete measure γ and let $\tilde{G} := G + c$ $(c \neq 0)$ be another d.f. of the same measure γ. With φ increasing and absolutely continuous, put $H := \varphi \circ G$ and $\tilde{H} := \varphi \circ \tilde{G}$. Then, in general, the measure θ determined by H and the measure $\tilde{\theta}$ determined by $\tilde{H}$ are *two different measures*. For example, let γ be the unit mass at 0 and $\varphi(x) = e^x$. Then $G = \mathbf{I}_{[0,\infty)}$ and $\tilde{G} = \mathbf{I}_{[0,\infty)} + 1$ are d.f.'s of γ, the measures θ and $\tilde{\theta}$ are concentrated on $\{0\}$, and

$$\theta\{0\} = \Delta H(0) = e^1 - e^0 \qquad \text{whereas } \tilde{\theta}\{0\} = \Delta \tilde{H}(0) = e^2 - e^1.$$

As a concluding example, consider again the integral $\int_A f\,de^G$. It was already noted that $\int_A f\,de^G \neq \int_A f \cdot e^G\,dG$ is possible when G is not continuous. In fact, if G is a discrete d.f. concentrated on $\mathscr{M}$, it follows from Theorem (6) that (in the usual sense)

$$\int_A f\,de^G = \sum_{x \in A \cap \mathscr{M}} f(x)\{e^{G(x)} - e^{G(x^-)}\} = \sum_{x \in A \cap \mathscr{M}} f(x) e^{G(x^-)}\{e^{\Delta G(x)} - 1\}.$$

REFERENCES

1. P. Billingsley, "Probability and Measure," 2nd ed., Wiley, New York, 1986.
2. E. Hewitt and K. Stromberg, "Real and Abstract Analysis," Springer-Verlag, New York/Berlin, 1965.
3. E. Kamke, "Das Lebesgue-Stieltjes Integral," Teubner, Leipzig, 1956.
4. P. E. Kopp, "Martingales and Stochastic Integrals," Cambridge Univ. Press, Cambridge, UK, 1984.
5. H. Lebesgue, "Leçons sur l'Intégration," 2nd ed., Gauthiers–Villars, Paris, 1928; reprint, Chelsea, New York, 1973.
6. R. S. Liptser and A. N. Shiryayev, "Statistics of Random Processes," Vol. I, Springer-Verlag, New York/Berlin, 1978.
7. E. Parzen, Nonparametric statistical data modeling, *J. Amer. Statist. Assoc.* **74** (1979), 105–121.
8. F. Riesz and B. Sz.-Nagy, "Functional Analysis," Ungar, New York, 1955.
9. H. L. Royden, "Real Analysis," 2nd ed., Macmillan, Co., New York, 1968.
10. W. Rudin, "Real and Complex Analysis," 3rd ed., McGraw–Hill, New York, 1987.
11. S. Saks, "Theory of the Integral," 2nd ed., Warsaw, 1937; reprint, Dover, New York, 1964.
12. B. B. Winter, "On Lebesgue-Stieltjes Integration and Some Related Topics," Laboratory for Research in Statistics and Probability, Ottawa, 1992.